Periodicals of Engineering and Natural Sciences ISSN 2303-4521
Vol. 8, No. 2, June 2020, pp.673-683

Studying some approaches to estimate the smoothing parameter for the nonparametric regression model

Mohammed Saffaa Wanas [1], Mohammed H.AL-Sharoat [2]
[1,2] Statistics Department, AL Qadisiyah University

ABSTRACT

Several previous studies have addressed various topics in regression analysis and estimation of the appropriate regression equation. It assumes that there is a known and pre-defined function relationship between variables. The studied variables are known for distribution using some known methods of estimation, such as the ordinary least squares method (OLS) and the maximum likelihood method (MLE). The parameter model can be estimated due to problems arising from the application of the parameter model, because the theoretical assumptions of the model application are not met. Here, we adopted another method of estimating the regression equation using non-parametric methods. It proved its efficiency and ability to analyze data without the need for prior assumptions on the model. Based on the adopted data, it determines the functional shape of the studied population. Therefore, the aim of this research is to use non-parametric smoothing methods to approximate the non-parametric regression function to the real regression function. This is done by using some non-parametric smoothing methods such as Kernel methods by Nadaraya-Watson and the method of the nearest neighbor (K-Nearest-Neighbor) depending on the bandwidth (h).The study uses the experimental method of simulation on two test functions. Three sizes of sample data (n = 15, n = 50, n = 75) and three values for random error variance $(\sigma^2 = 0.5), (\sigma^2 = 1), (\sigma^2 = 2)$ are assumed. Kernel methods based on Nadaraya-Watson Smoothed Cross Validation are the best choice for the bandwidth of the first test function. On the other hand, Least Squared Cross Validation method for the forensic crossing is the best choice for the bandwidth of the second test function. The second one was better than the neighbor method closest to the first test function.

Keywords: Nadaraya-Watson, K-Nearest-Neighbor, Least Squared Cross Validation, Biased Cross Validation

Corresponding Author:

Mohammed Saffaa Wanas
Department of Statistics
AL- Qadisiyah University
College of Administration &Economics
Email: malsharood@gmail.com

1. Introduction

The enormous potential and rapid development of electronic computers in accomplishing complex calculations have made non-parametric statistical methods more attractive to researchers than methods of estimating the regression of teachers. This is because of their high flexibility and lack of rigidity in computerized statistical methods [1]. The imposition of severe restrictions on the model and the researcher's interest in non-parametric regression models is to give a general description of the relationship between the explanatory variable X_i and the response variable Y_i and does not study the details of that relationship. Therefore, the non-parametric Kernel function by an estimator (Nadaraya-Watson) will be used to find the estimated regression. The estimate of m in the regression model in equation (1) is as follows [2]:

$$Y_i = m(X_i) + \epsilon_i \qquad i = 1, \dots, n \qquad (1)$$

The problem of applying parameter regression models occurs when the specific assumptions for applying this model are not met. Therefore, the researchers use the non-parameter regression model, which does not impose limitations on the model.

The aim of this study is to estimate the bandwidth of the non-parametric regression model of the Kernel function using the Nadaraya Watson smoothing and the Nearest neighbor method with parametric analyses.

2. Kernel smoothing (Nadaraya-Watson)

It is a non-parametric method proposed by Nadaraya and Watson in 1964 [2]. They are the first to use this estimator based on the method of serial weights. This method estimates the function m in equation (1), and it is characterized as being without parameter. This estimator is characterized by having a specific and continuous function and its integration is equal to unity . The general formula of Nadaraya-Watson estimator is as follows:

$$\widehat{m}_h(x) = \frac{\sum_{i=1}^{n} K_h(x - X_i) y_i}{\sum_{i=1}^{n} K_h(x - X_i)} \quad \ldots (2)$$

$$\because \; W_h(x - X_i) = \frac{K_h(x - X_i)}{\sum_{i=1}^{n} K_h(x - X_i)} \quad \ldots (3)$$

$$\therefore \; \widehat{m}_h(x) = \sum_{i=1}^{n} W_h(x - X_i)\, y_i \quad \ldots (4)$$

Where,
$W:$: Weight function
$h:$ Bandwidth
$K_h(x - X_i)$: The function kernel

$$\sum_{i=1}^{n} W_h(x - X_i) = 1 \quad \ldots (5)$$

Kernel functions generally have the following characteristics [3]:

- K (u) is a specific probability density function (non-negative) in the case of K (u) ≥0 for each (u), such that K (u) :R → R, as well as having symmetric characteristics.
- Moments for Kernel function are calculated as follows:

$$M_j(K) = \int_{-\infty}^{\infty} u^j\, K(u) d(u) \quad \ldots (6)$$

- Symmetry K (u) = K (-u) for each (u).
- All individual moments are equal to zero.

$$\int_{-\infty}^{\infty} u\, K(u) d(u) = 0 \quad \ldots (7)$$

$$C_K = R(K) = \int_{-\infty}^{\infty} K^2(u) d(u) \quad \ldots (8)$$

$$d_K = M_2(K) = \int_{-\infty}^{\infty} u^2 K(u) d(u) \quad \ldots (9)$$

Where, k represents the degree of Kernel (Kernel order).
Some of the commonly used Kernel functions can be summarized as shown in Table 1:

Table 1. Kernel functions

Kernel	K(u)	
Gaussian	$k(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right)$	I(-∞, ∞)

Kernel	K(u)	
Uniform	$k(u) = \frac{1}{2}$	[-1, 1]
Epanechnikov	$k(u) = \frac{3}{4}(1 - u^2)$	[-1, 1]
Triweight	$k(u) = \frac{35}{32}(1 - u^2)^3$	[-1, 1]

3. Estimation methods

The Kernel method was used by Nadaraya-Watson smoothing to find the estimated regression equation. This estimator depends on bandwidth (h) which can be calculated based on the following:

1. Least Squared Cross Validation method.
2. Biased Cross Validation method.
3. Smoothed Cross Validation method.

3.1. Least squared cross validation method (LSCV)

This method proposed Bowman in 1984 , is the most widely used method, and it is one of the best studied methods is to estimate $\widehat{\mathrm{h}}_{\mathrm{o}}$ from minimizing the integrated square error (ISE h) of the bandwidth [4,5] .

$$\mathrm{ISE(h)} = \int_{-\infty}^{\infty} \left[\hat{\mathrm{f}}(\mathrm{x}) - \mathrm{f}(\mathrm{x})\right]^2 \mathrm{dx}$$

$$= \int_{-\infty}^{\infty} \hat{\mathrm{f}}^2(\mathrm{x})\mathrm{dx} - 2\int_{-\infty}^{\infty} \hat{\mathrm{f}}(\mathrm{x})\mathrm{f}(\mathrm{x})\mathrm{dx} + \int_{-\infty}^{\infty} \mathrm{f}^2(\mathrm{x})\mathrm{dx} \ldots (10)$$

Note that:

$\int_{-\infty}^{\infty} \mathrm{f}^2(\mathrm{x})\mathrm{dx}$: The amount does not depend on the bandwidth h

$\int_{-\infty}^{\infty} \hat{\mathrm{f}}^2(\mathrm{x})\mathrm{dx}$: Information value calculated from data

$\int_{-\infty}^{\infty} \hat{\mathrm{f}}(\mathrm{x})\mathrm{f}(\mathrm{x})\mathrm{dx}$: The only amount to be estimated.

By subtracting the constant, we can perceive that the reduction of ISE relative to h is equal to:

$$\mathrm{ISE(h)} - \int_{-\infty}^{\infty} \mathrm{f}^2(\mathrm{x})\mathrm{dx} = \int_{-\infty}^{\infty} \hat{\mathrm{f}}^2(\mathrm{x})\mathrm{dx} - 2\int_{-\infty}^{\infty} \hat{\mathrm{f}}(\mathrm{x})\mathrm{f}(\mathrm{x})\mathrm{dx} \ldots (11)$$

Leave-one-out Cross Validation was used to estimate $\int_{-\infty}^{\infty} \hat{\mathrm{f}}(\mathrm{x})\mathrm{f}(\mathrm{x})\mathrm{dx}$, which represents E $\breve{f}(x)$ and it is equal to:

$$\mathrm{E}\,\hat{\mathrm{f}}(\mathrm{x}) = \mathrm{n}^{-1}\sum_{\mathrm{i}=1}^{\mathrm{n}} \hat{\mathrm{f}}_{-\mathrm{i}}(\mathrm{X}_{\mathrm{i}}) \qquad \dots(12)$$

From the offset the value $\mathrm{E}\,\hat{\mathrm{f}}(\mathrm{x})$, We get LSCV as follows:

$$\mathrm{LSCV}(\mathrm{h}) = \int_{-\infty}^{\infty} \hat{\mathrm{f}}^2(\mathrm{x})\mathrm{dx} - \frac{2}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}} \hat{\mathrm{f}}_{-\mathrm{i}}(\mathrm{X}_{\mathrm{i}}) \qquad \dots(13)$$

This method can also be called an unbiased forensic transit method. It is introductory parameter that reduces the function:

$$\hat{h}_{LSCV} = argmin_h LSCV(h) \qquad \dots(14)$$

3.2. Biased cross validation method (BCV)

This method was proposed by Scott and Terrell (1987), which is based on the average integrated square error [4, 6]:

$$AMISE(h) = (nh)^{-1}R(K) + h^4\left(\frac{M_2(K)}{2}\right)^2 R(f^2)$$
$$= \frac{R(K)}{nh} + \frac{h^4 d_K^2}{4} R\left(\acute{\hat{f}}\right) \qquad \dots(15)$$
$$\because R\left(\acute{\hat{f}}\right) = R(f^2) = \int_{-\infty}^{\infty} \left[\acute{\hat{f}}(x)\right]^2 dx$$

The BCV (h) function is obtained by replacing the unknown values in $R\left(\acute{\hat{f}}\right)$ with:

$$\tilde{R}\left(\acute{\hat{f}}\right) = R\left(\acute{\acute{\hat{f}}}\ (.,h)\right) - \frac{R\left(\acute{\hat{K}}\right)}{nh^5} = n^{-2}\sum_{i\neq j}\left(\acute{\hat{K}}_h * \acute{\hat{K}}_h\right)(X_i - X_j) \quad \dots(16)$$

The introductory parameter that reduces the function is:

$$\hat{\mathrm{h}}_{\mathrm{BCV}} = \mathrm{argmin}_{\mathrm{h}\in\mathrm{H}_{\mathrm{n}}}[\mathrm{BCV}(\mathrm{h})] \qquad \dots(17)$$

3.3. Smoothed cross validation method (SCV)

This method was suggested by Park and Marron in 1992, which is based on the mean integrated square error MISE (h) [6, 7]:

$$MISE(h) = IV(h) + IB(h)$$

Where, IV is integrated variance and IB is integrated squared bias. They can computed by:

$$IV(h) = \int_{-\infty}^{\infty} var\left(\hat{f}(x)\right)dx = \frac{C_K}{nh} + \frac{1}{n}\int_{-\infty}^{\infty}(K_h * f)^2(x)dx$$
$$IB(h) = \int_{-\infty}^{\infty} bias^2\left(\hat{f}(x)\right)dx = \int_{-\infty}^{\infty}(K_h * f - f)^2(x)dx$$
$$= \int_{-\infty}^{\infty}(K_h * f)^2(x)dx - 2\int_{-\infty}^{\infty}(K_h * f)(x)f(x)dx + \int_{-\infty}^{\infty} f^2(x)dx \qquad \dots(18)$$

The SCV is originated from LSCV as n≈n-1. If we can use the Leave-one-out of the experimental estimator $\left(\hat{f}_{P,-i}(x;g)\right)$, we can use the following equation:

$$LSCV = \frac{C_K}{nh} + \frac{1}{n(n-1)}\sum\sum_{i\neq j}(K_h * K_h - 2K_h)(X_i - X_j)$$

$$= \frac{C_K}{nh} + \frac{1}{n(n-1)} \sum \sum_{i \neq j} (K_h * K_h - 2K_h + K_0)(X_i - X_j) \quad \dots (19)$$

The last equation is obtained if there is no repetition of data as in one probability of continuous data. SCV (h) is equal to LSCV (h) at g = 0. They also suggested that $\hat{h}_{SCV} = \hat{h}_{SCV}(g)$, which comes from the reduction of SCV (h):

$$\hat{h}_{SCV} = \operatorname{argmin}_{h \in H_n} [SCV(h)] \quad \dots (22)$$

Their results also showed that the relative convergence rate $\hat{h}_{SCV}$ for $h_0$is $O_P(n^{-1/2})$ as the best ratio achieved.

4. Nearest neighbour method

Random model can be resultant from the ratio of the average real distance between each adjacent site in the region to the average distance between the same numbers of sites if they were distributed randomly in the same area. If the bandwidth is adjusted to include a fixed number of observations (k views), the estimated k-Nearest Neighbor (k-NN) is estimated for point x. The scale is based on knowing the distance between each location representing the point in the area and adjacent locations. The actual distance stands for the sum of the actual distances divided by the number of distances which is always equal to the number of sites. It can be compared it with the average expected distance (random distance) in the theoretical random distribution [8].
The Kernel function W_{ni}(x) can be expressed as:

$$W_{ni}(x) = \frac{k(\frac{xi-xj}{kn})}{\sum_{j=1}^{n} k(\frac{xi-xj}{kn})} \quad (23)$$

$$i, j = 1,2, \dots n$$

Since k (.) represents a restricted and non-negative kernel function, and u> 1 is achieved for all values of k (u) = 0, *Kn* represents the Euclidian distance between x and k from the nearest neighbor of x. Here, *K* = *Kn* with kn⟶∞ and n⟶∞. In the other words, k represents the bandwidth similar to the h value in Nadaraya- Watson estimator if the Kernel functions are applied to this estimator as in the Nadaraya- Watson model. The introductory parameter k is calculated by:

$$Kn = d(xi,xj) = \sqrt{\sum_{k=1}^{n} (xik - xjk)^2} \quad ,i,j = 1,2,\dots n \qquad (24)$$

K-NN estimator is heavily influenced by the boot parameter K. The K parameter adjusts the degree of Smoothing of the estimated curve. By compensating for a value in equation (2), we get the estimate of the nearest neighbor:

$$\hat{m}(X) = \frac{\sum_{I=1}^{n} k(\frac{xi-xj}{kn}) yi}{\sum_{j=1}^{n} k(\frac{xi-xj}{kn})} \quad , i, j = 1,2, \dots n \qquad (25)$$

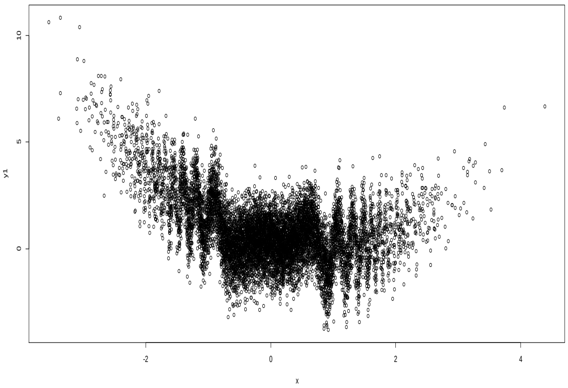
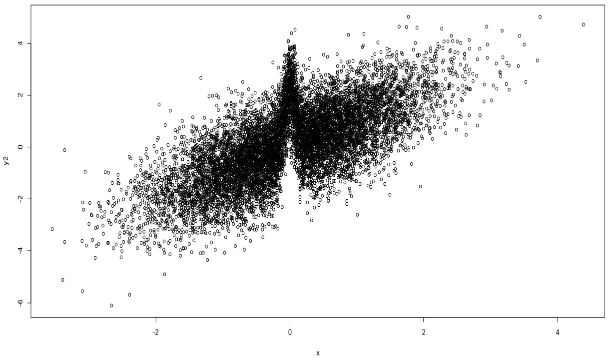
5. The simulation

Simulation was used to generate the observations of the study according to the sample sizes (n = 15,50, and 75). The random error is distributed according to the standard normal distribution with an average equal to zero and variance. The Ndaraya-Watson methods and Nearest Neighbor mentioned in the theoretical aspect and for all sample sizes were applied. AMSE standard was adopted in the differentiation between the estimation methods according to the following steps:

1- Generating the data of the explanatory variables (X_i), so that they are distributed in a standard normal distribution, X_i~N(0,1).

2- Two test functions were selected as follows:

Table 2. The functions used in the simulation

Curved shape	Formula	Test function
	**Y(x)= sin(2$\boldsymbol{\pi x^3}$) + $\mathbf{0.5}(\boldsymbol{x}-\mathbf{0.75})^{\mathbf{2}}$**	Nonlinear function
	**Y(x)=x+2exp(-64$\boldsymbol{x^2}$)**	Nonlinear function

3- Generate random errors so that a standard normal distribution was distributed with an average of zero and 1 variation;
ei ~N(0,1) ,i=1,2,…,n
4-Dependent Variable: The adopted variable is calculated by adding the functions of the explanatory variable m (x) plus the random error. Three sample sizes were selected (n = 15, 50, and 75). Three levels of variance were used:
.$\sigma^2 = 2$ and $\sigma^2 = 1$ and $\sigma^2 = 0.5$
5-The simulation experiments were relied on a kind of Kernel function as a weight function in the kernel methods (Epanechnikov).

6-Simulation experiment was repeated for each case 1000 times to ensure randomization.

7-The methods of estimation by Nadaraya- Watson smoothing are compared based on the standard accuracy rate of average mean square error (AMSE), to choose the best bandwidth. The results are shown in Table 3.

Table 3. The results of the AMSE standard for cross validation methods for the selection of the N-W bandwidth of the first test function with different sample sizes and variance

The method	Sample size	$\sigma^2 = 0.5$	$\sigma^2 = 1$	$\sigma^2 = 2$
LSCV Bandwidth	15	0.019655782	0.02145935	0.02871872

	50	0.018795010	0.02070371	0.02514760
	75	0.18918134	0.02025397	0.02581250
BCV Bandwidth	15	0.01996442	0.02110338	0.02718635
	50	0.019649987	0.02067132	0.02303550
	75	0.019920357	0.02080189	0.02444422
SCV Bandwidth	15	0.01923976	0.02151745	0.03053390
	50	0.018002194	0.02091579	0.02646993
	75	0.018218113	0.02008344	0.02780044

6. Analysis of the results

Through Table 3 of the first test function, we found that:
Y(x)= $\sin(2\pi x^3) + 0.5(x - 0.75)^2$

At the variance level $\sigma^2 = 0.5$, the results have shown the following:

1- For the sample size (n = 15), Kernel (h: SCV) preference was found where the value of the differential criterion was AMSE = 0.01923976, then the Kernel estimator (h: LSCV) with the value of the differential criterion AMSE = 0.019655782 and the Kernel estimator (h: BCV) with the criterion value. The trade-off is in the case of AMSE = 0.01996442.
2- When increasing the sample size (n = 50), we notice that the Kernel estimator (h: SCV) is given a preference: AMSE = 0.018002194, then Kernel (h: LSCV) with the value of AMSE = 0.018795010, followed by the Kernel (h: BCV) estimator. The trade-off is in the case of AMSE = 0.019649987.
3-By increasing the sample size (n = 75), we notice that the Kernel estimator (h: SCV) is given a differential criterion in the case of AMSE =0.018218113 and Kernel (h: LSCV) with a differential criterion of AMSE =0.018918134. Finally, it is followed by the Kernel estimator (h: BCV) with the differential criterion in the case of AMSE =0.019920357.
At the variance level $\sigma^2 = 1$, the results have shown the following:

1- For the sample size (n = 15), the preference for Kernel estimator (h: BCV) was found. The value of the differential criterion was AMSE = 0.02110338, then the Kernel estimator (h: LSCV) with the value of differentiation criterion in the case of AMSE= 0.02145935 and the Kernel estimator (h: SCV) with the criterion value. The trade-off is in the case of AMSE = 0.02151745.

2- For the sample size (n = 50), the preference for Kernel estimator (h: BCV) was shown in the case of AMSE = 0.02067132, Kernel (h: LSCV) with AMSE=0.02070371, and Kernel (h: SCV) with the criterion value. The trade-off is in the case of AMSE = 0.02091579.
3-For the sample size (n = 75), the preference for Kernel estimator (h:SCV) was shown in the case of differential criterion is AMSE =0.02008344. Then, Kernel estimator (h: LSCV) under differential criterion is with AMSE=0.02025397. Kernel estimator (h: BCV) with differential criterion is with AMSE=0.02080189.

At the variance level $\sigma^2 = 2$, the results have shown the following:
1-For the sample size (n = 15), the preference for Kernel estimator (h: BCV) was shown where the value of the differentiation criterion was AMSE = 0.02718635. Then, Kernel estimator (h: LSCV) is with the value of the

differentiation criterion = 0.02871872 = AMSE. Kernel estimator (h: SCV) under differentiation criterion is with AMSE = 0.03053390.

2- For sample size (n = 50), Kernel estimator (h: BCV) was shown to have a preference value of AMSE = 0.02303550. Then, Kernel (h: LSCV) is with AMSE=0.02514760 and Kernel (h: SCV) is in the case of AMSE differential criterion = 0.02646993.

3- For the sample size (n = 75), the preference for Kernel estimator (h: BCV) is in the case of the differential criterion with AMSE =0.02444422. Then, the Kernel estimator (h: LSCV) is with the value of the differential criterion AMSE=0.02581250. Kernel estimator (h: SCV) under criterion value with trade-off of AMSE =0.02780044.

Table 4. The results of the AMSE criterion for selecting the bandwidth h in the manner of the nearest neighbor (Nearest- Neighbor) of the first test function with the sample sizes and different variances

The method	Sample size	$\sigma^2 = 0.5$	$\sigma^2 = 1$	$\sigma^2 = 2$
k-Nearest Neighbor	15	0.008795535	0.01611161	0.04562406
	50	0.008138867	0.01731858	0.04814688
	75	0.007397969	0.01632456	0.04543678

Through Table 4 of the first test function, we found that:
Y(x)= sin($2\pi x^3$) $+ 0.5(x - 0.75)^2$

At the variance level $\sigma^2 = 0.5$, the preference of the closest neighbor method to the sample size was 75 based on the value of the differentiation criterion AMSE=0.007397969. This is followed by the neighbor method closest to the size of the curse 50 based on the value of the criterion of trade-off with AMSE=0.008138867. Finally, the neighbor method came very close to the sample size of 15 by value of AMSE=0.008795535.
At the variance level $\sigma^2 = 1$, the preference of the closest neighbor method to the sample size was 15 based on the value of the differentiation criterion AMSE=0.01611161 This is followed by the neighbor method closest to the size of the curse 75 based on the value of the criterion of trade-off AMSE=0.01632456. Finally, the neighbor method came very close the sample size of 50 by AMSE=0.01731858. At the variance level $\sigma^2 = 2$, the preference of the closest neighbor method to the sample size was 75 based on the value of the differentiation criterion with AMSE=0.04543678. This is followed by the neighbor method closest to the size of the curse 15 based on the value of the criterion of trade-off with AMSE=0.04562406. Finally, the neighbor method came very close to the sample size of 50 by AMSE=0.04814688.

Table 5. The results of the AMSE standard for cross validation method for the selection of the N-W bandwidth of the second test function with different sample sizes and variances

The method	Sample size	$\sigma^2 = 0.5$	$\sigma^2 = 1$	$\sigma^2 = 2$
LSCV Bandwidth	15	0.009967032	0.02398877	0.05370734
	50	0.010147650	0.01859501	0.05584223
	75	0.011036842	0.02058916	0.05806116
BCV Bandwidth	15	0.01018051	0.02476190	0.05334866

	50	0.010286902	0.01874658	0.05589809
	75	0.011174065	0.02035461	0.05716124
	15	0.01034695	0.02236297	0.05342132
SCV Bandwidth	50	0.009917725	0.01826048	0.05663094
	75	0.011412566	0.02146697	0.06161511

Through Table 5 of the second test function, we found that:

Y(x)=x+2exp(-64x^2)

At the variance level $\sigma^2 = 0.5$, the results has shown the following:

1- For the sample size (n = 15), the Kernel estimator (h: LSCV) was given preference, where the value of the differential criterion was AMSE = 0.009967032. Then, Kernel estimator (h: BCV) under differentiation criterion was with AMSE=0.01018051.The Kernel estimator (h: SCV) with a criterion value under trade-off was with AMSE = 0.01034695.

2- When increasing the sample size (n = 50), we notice that the Kernel estimator (h: SCV) has a preference: AMSE = 0.009917725. Then, Kernel (h: LSCV) is with the value of AMSE = 0.010147650, followed by the Kernel (h: BCV) estimator. The trade-off is AMSE = 0.010286902.

3- When increasing the sample size (n = 75), we notice the preference of the Kernel estimator (h: LSCV). The differentiation criterion is given by AMSE =0.011036842. Then, Kernel estimator (h: BCV) is given by the differential criterion with AMSE = 0.011174065, followed by Kernel estimator (h: SCV) under differential criterion with AMSE = 0.011412566.

At the variance level $\sigma^2 = 1$, the results showed:

1- For sample size (n = 15), Kernel (h: SCV) preference was shown where the value of the differential criterion was AMSE = 0.02236297. Kernel estimator (h: LSCV) under differential criterion is with AMSE 0.02398877 and the Kernel estimator (h: BCV) under a criterion value under trade-off is with AMSE = 0.02476190.

2- For the sample size (n = 50) the preference for Kernel estimator (h: SCV) has been shown as the value of the differentiation criterion is with AMSE = 0.01826048, then the Kernel estimator (h: LSCV) with the value of differentiation criterion = 0.01859501 = AMSE. Kernel estimator (h: BCV) with the criterion value under trade-off is with AMSE = 0.01874658.

3- For the sample size (n = 75), the preference for Kernel estimator (h: BCV) was shown where the value of the differential criterion was with AMSE =0.02035461. Kernel estimator (h: LSCV) under differentiation criterion has been with AMSE=0.02058916, while Kernel estimator (h: SCV) under differential criterion is with AMSE = 0.02146697.

At the variance level $\sigma^2 = 2$, the results showed the following:

1- For the sample size (n = 15), Kernel estimator (h: BCV) was favored with the AMSE differential value of 0.05334866. Kernel (h: SCV) estimator has been with differential AMSE of 0.05342132, and the Kernel (h: LSCV) estimator under trade-off is with AMSE = 0.05370734.

2- For the sample size (n = 50), the preference for Kernel estimator (h: LSCV) was shown where the value of the differential criterion was with AMSE = 0.05584223. Kernel estimator (h: BCV) under differential criterion was with AMSE=0.05589809. Kernel estimator (h: SCV) with AMSE differential criterion has been equal to 0.05663094.

3-For the sample size (n = 75), the preference for Kernel estimator (h: BCV) was shown where the value of the differential criterion was with AMSE =0.05716124. Kernel estimator (h: LSCV) was with the value of the differentiation criterion of AMSE=0.05806116. Kernel estimator (h: SCV) under differentiation criterion is with AMSE =0.06161511.

Table 6. The results of the AMSE criterion for selecting the bandwidth h in the manner of the nearest neighbor (Nearest- Neighbor) of the second test function with the sample sizes and different variance

The method	Sample size	$\sigma^2 = 0.5$	$\sigma^2 = 1$	$\sigma^2 = 2$
k-Nearest Neighbor	15	0.005635798	0.01274303	0.04101418
	50	0.007675205	0.01290641	0.04769047
	75	0.007045684	0.01403334	0.04782431

Through Table 6 of the second test function, we found
Y(x)=x+2exp(-64x^2)
At the variance level $\sigma^2 = 0.5$, the preference of the closest neighbor method to the sample size was 15 based on the value of the differentiation criterion with AMSE=0.005635798. This is followed by the neighbor method nearby the size of the curse of 75 based on the value of the criterion of trade-off with AMSE=0.007045684. Finally, the neighbor method has been very close to the sample size of 50 by AMSE=0.007675205.
At the variance level $\sigma^2 = 1$, the preference of the closest neighbor method to the sample size was 15 based on the value of the differentiation criterion with AMSE=0.01274303.This is followed by the neighbor method closest to the size of the curse 50 based on the value of the criterion of trade-off AMSE=0.01290641 finally came the neighbor method closest to the sample size of 75 by value AMSE=0.01403334.
At the variance level $\sigma^2 = 2$, the preference of the closest neighbor method to the sample size was 15 based on the value of the differentiation criterion AMSE=0.04101418 This is followed by the neighbor method closest to the size of the curse of 50 based on the value of the criterion of trade-off with AMSE=0.04769047. Finally the neighbor method has been very close to the sample size of 75 by AMSE=0.04782431.

7. Conclusions

Smoothed Cross-Validation is the best non-parameter smoothing for the first test of sample sizes (n = 15,50, 75) with a variance level of $\sigma^2 = 0.5$. Biased Cross Validation based on sample size (n = 15,50) under levels of variance of $\sigma^2 = 1$ and 2 has been the finest method used. The best nonparameterization methods for the second test function is under sample sizes (n = 15, 75) with level of variance of $\sigma^2 = 0.5$. The method of Smoothed Cross-Validation for the sample size (n = 15,50) and the level of variance $\sigma^2 = 1$, has given the best results. For the level of variation with $\sigma^2 = 2$, Biased Cross Validation method is the best technique.
The nearest neighbor method of the second test function has better results than the nearest neighbor method of the first test function under employed sample sizes.

References

[1] M. Hanif, "Reweighted Nadaraya-Watson estimator of scalar diffusion models by using asymmetric kernels," Far East Journal of Psychology and Business, vol. 4, no.5 ,pp. 53-69,2011.

[2] W. Nadarajah, "Some new estimates for distribution functions," Theory Probability Appl., vol.15, no. 3, pp.497-500, 1964

[3] L. Molanes and R. Cao, "Plug-in bandwidth selector for the kernel relative density estimator", *Annals of the Institute of Statistical Mathematics AISM*, vol.60, no.2, pp. 273-300, 2008

[4] R. Abdel and J. Jetter, "A simulation study for the bandwidth selection in the kernel density estimation based on the exact and the asymptotic mise," *Pak. J. Statist*, vol. 26, no. 1, pp. 239-265, 2010.

[5] A. Bowman, "An Alternative Method of Cross-Validation for the Smoothing of Density Estimates," *Biometrika*, vol. 71, no. 2, pp. 353-360, 1984

[6] T. Duong and M. Hazelton, "Cross-validation Bandwidth Matrices for Multivariate Kernel Density Estimation," *Board of the Foundation of the Scandinavian Journal of Statistics, Blackwell Publishing Ltd*, vol. 32, no. 3, pp. 485–506, 2005.

[7] B. Park and S. Marron, "On the Use of Pilot Estimators in Bandwidth Selection, Nonparametric Statistics", *Seoul national university-and university sity of North Carolina*, vol. 1, no.3, PP. 231-240, 1992.
[8] S. Noor, "Estimation of non-teacher regression functions using some methods of boot," *Iraqi Journal of Statistical Sciences*, vol. 20, no. 20, pp. 393-373, 2011.